First Year Annual Report
(For Period: 11/16/2005 - 09/01/2006)

Automated Classification of X-ray Sources for Very Large Datasets

AISRP Grant Number: NNG04GQ07G

PI: Susan Hojnacki
Co-I: Joel Kastner
Collaborators: Giusi Micela, Steven LaLonde

The initial period of performance for this AISRP grant was November 1, 2004 thru October 31, 2006. Research on this grant began in August of 2005, after the PI was hired at the Rochester Institute of Technology. A 12-month extension (through Oct. 31, 2007) has been granted for this project.

INTRODUCTION

A large fraction of Chandra X-ray Observatory (*Chandra*) and XMM-Newton observing time has been devoted to the study of young stellar clusters and, as a consequence, large datasets exist from these observations of rich stellar fields. A typical *Chandra* or XMM-Newton CCD observation of a young stellar cluster results in detection of X-ray emissions from tens to hundreds of very young stars. Spectral and temporal data is available from these observations. A wide range of temporal behavior has been detected in X-ray sources. This temporal behavior ranges from long-term flaring to episodic, short X-ray bursts.

Our objective is to develop an algorithm that uses both spectral data and temporal properties to group and classify discrete sources of X-ray emission. This new method will handle large quantities of data and operate independently of the requirement of source models (spectral or temporal) and a priori knowledge concerning the nature of the sources (i.e., young stars, interacting binaries, active galactic nuclei). This unbiased approach could lead to the discovery of new classes of sources that do not fit any existing models and X-ray sources that are extreme outliers in the spectral and/or temporal domains.

BACKGROUND

Our spectral classification algorithm groups X-ray sources based on their X-ray spectral properties and can handle large quantities of data (Hojnacki & Kastner 2004). It operates independently of the requirement of source models and a priori knowledge concerning the nature of the sources (e.g., young stars, interacting binaries, active galactic nuclei). The Chandra Orion Ultradeep Project (COUP) observation of the Orion Nebula Cluster (ONC), an ~838 ks exposure obtained over a 13-day period in January of 2003, represents the most sensitive and comprehensive description of X-ray emission from a young star cluster (Getman et al. 2005). Over 1600 X-ray emitting young stars have been detected in the COUP observation (Getman et al. 2005). Our analysis of the COUP data resulted in 17 X-ray spectral classes of sources from a sample of 444 X-ray-emitting objects in the ONC. A plot of the values of the first two principal components (PCs) for each source is shown in Figure 1. Average spectra of some of the classes are shown. The progression of classes moving clockwise around the horseshoe-shaped plot forms a sequence of decreasing spectral hardness. We found strong similarities between spectra of a given class, whereas fundamental spectral differences were found between different classes. The resulting X-ray spectral classes were compared with known optical and near-infrared (ONIR) properties of the young stellar population of the ONC. In Table 1, we list the mean hydrogen column density (N_H), mean effective (photospheric) temperature (T_{eff}), visual extinction (A_V), and mean near-infrared excess ($\Delta(I - K)$) of the ONIR counterparts of the members of a subset of the X-ray classes. Values in parentheses represent error on the mean. These results were compiled from data available for the X-ray-emitting ONC population (Getman et al. 2005).

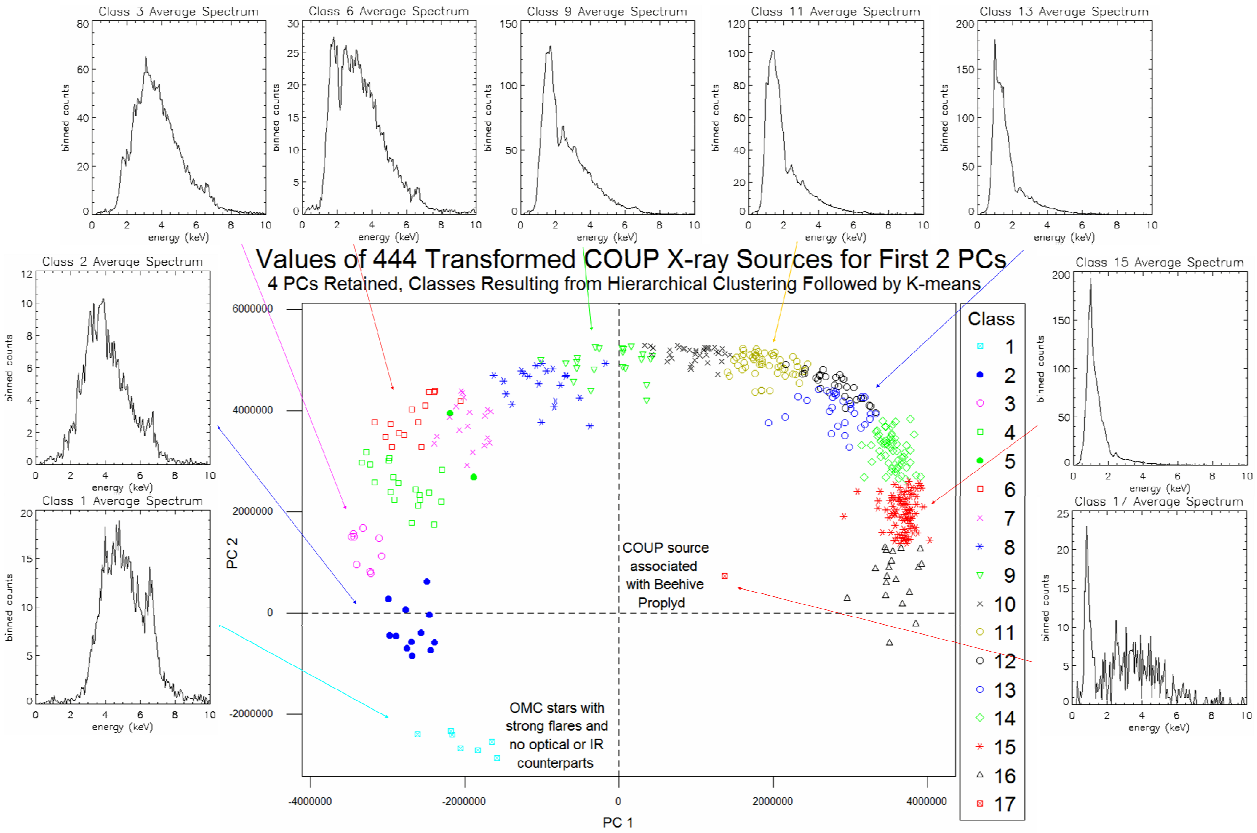

Figure 1: Plot of the first 2 principal components showing the source classes. The class numbers increase clockwise around the horseshoe. Average spectra for some of the classes are shown around the outside of the PCA plot.

Using our algorithm on the COUP sample, trends have been found between X-ray spectral parameters and stellar parameters for very low-mass, soft spectra, young sources. Specifically, the hydrogen column density decreases monotonically from class 1 to class 16. The large log N_H value for classes 1 through 8 results in small fractions of ONIR counterparts within those classes. For classes 11 through 15, which have relatively large numbers of ONIR counterparts, our analysis reveals an apparent trend of increasing spectral softness with decreasing T_{eff}. In addition, the near infrared excess is observed to decrease monotonically from class 10 to class 16, suggesting a generally decreasing accretion rate.

Table 1: Mean values of COUP X-ray spectral classes.

Class	# of Sources	<log N_H> atoms/cm^2	N	<log Teff> K	N	A_V mag	N	Δ(I-K) mag	N
1	7	23.4 (0.06)	7		0		0		0
2	12	22.96 (0.03)	12	3.57:	1		0		0
3	9	22.79 (0.02)	9		0		0		0
4	19	22.66 (0.01)	19	3.59:	3	1.34:	3	1.66:	3
5	2	22.52 (0.05)	2	3.68:	1	3.67:	1	2.61:	1
6	14	22.48 (0.02)	14		0		0		0
7	18	22.46 (0.02)	18	3.7:	1	3.52:	1	0.98:	1
8	21	22.3 (0.02)	21	3.55:	3	1.52:	3	0.3:	2
9	22	22.18 (0.01)	22	3.56 (0.02)	7	1.77 (0.99)	7	1.1:	3
10	37	22.03 (0.02)	37	3.58 (0.01)	21	2.6 (0.45)	20	1.31 (0.18)	18
11	54	21.9 (0.02)	54	3.57 (0.01)	38	2.69 (0.31)	38	0.91 (0.12)	30
12	30	21.66 (0.03)	30	3.59 (0.01)	20	1.57 (0.29)	19	0.8 (0.14)	16
13	30	21.61 (0.03)	30	3.56 (0.01)	22	1.44 (0.27)	22	0.62 (0.11)	18
14	61	21.32 (0.03)	61	3.55 (0.01)	45	1.16 (0.16)	44	0.49 (0.08)	38
15	88	20.79 (0.05)	86	3.52 (0.01)	75	0.65 (0.11)	72	0.25 (0.07)	62
16	19	20.28 (0.11)	19	3.50 (0.01)	14	0.24 (0.15)	14	0.11 (0.05)	12
17	1	20.88:	1	3.56:	1	0.34:	1		0

These trends show that the algorithm effectively sorts young stars into physically meaningful groups. Furthermore, the trends demonstrate that correlations exist between X-ray and ONIR properties of young stars in the ONC. These trends and correlations are of significance to the study of star formation and the mechanisms causing X-ray emission in young stellar clusters. One goal of our research is to determine if consistent correlations exist between X-ray spectral properties and ONIR attributes such as infrared excess, effective temperature, and/or other known stellar properties, similar to the correlations already found for the Orion region. To answer this question, we are currently examining additional star formation regions using our algorithm.

SUMMARY OF PROGRESS

During the first year of our project, we worked on the following areas:

- Verification of the algorithm using simulated X-ray sources
- Extension of the algorithm for use with other star formation regions
- Validation of ONC algorithm results by analysis of other star formation regions
- Exploration of input variables describing X-ray source variability

Results of our research in each of these areas are described in a series of papers now in preparation (see below) and are summarized as follows.

Simulation Results

The detailed response of Chandra to a variety of X-ray sources was simulated to aid in interpreting and verifying the X-ray spectral classes. Simulations of X-ray spectral data were run using XSPEC with a thermal plasma emission plus intervening absorption model. The log of the hydrogen column density (log N_H) was varied from 20.5 to 23.5 (atoms/cm^2) and the thermal plasma temperature (kT) was varied from 0.258 to 8.617 keV. This resulted in a total of 84 simulations. The 84 simulated sources were run through our X-ray spectral classification algorithm and plotted on the same axes as our subset of 444 COUP sources. The resulting 2-D principal component plot is shown in Figure 2. The 84 simulated sources are represented with solid diamonds and are labeled with their corresponding log N_H values. Sources of equal kT are connected by lines.

The distribution of simulated X-ray sources in PC-space reproduces the fundamental shape of the horseshoe-like plot formed by the COUP sample. Moving clockwise around the horseshoe, the simulations form a sequence of monotonically decreasing log N_H, from 23.5 in the extreme lower left of the plot to 20.5 in the lower right. This indicates that the position of a COUP source on the horseshoe is determined in large part by the degree of its obscuration by intervening cloud or circumstellar material. The importance of log N_H to the membership of the source classes is particularly apparent along the left side of the

horseshoe. In particular, the only simulations that can reproduce the position of the distinct grouping that forms class 1 are those for which log N_H = 23.5. It is also evident, however, that the positions of class 1 and class 2 sources cannot be reproduced by low-temperature plasma emission. At the other extreme, the high-numbered classes (14–16) appear to form a sequence of decreasing kT, indicating that plasma temperature is more important than intervening absorption in determining the positions of these sources along the horseshoe.

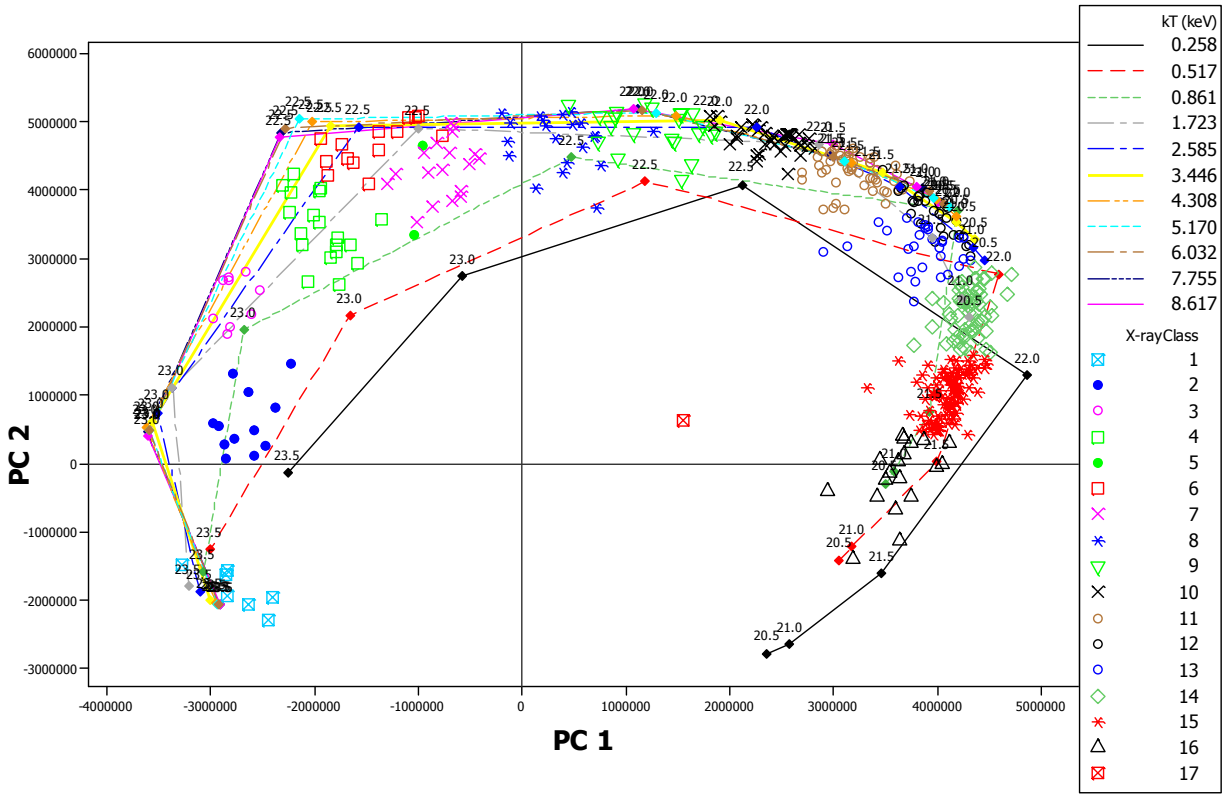

Figure 2: Results of 84 XSPEC simulations with varying kT and log N_H are overlaid on the principal component score plot of the 444 COUP sources. Lines represent kT. The points on the lines are the simulated sources and each is labeled with its corresponding log N_H.

Algorithm Extension for Use With Other Star Formation Regions

In order to classify X-ray sources in other star formation regions based on our current ONC source groupings, we have extended our existing X-ray spectral classification algorithm by

using discriminant analysis. This extended-algorithm is currently being tested and other methods may also be examined.

The extended algorithm consists of the following procedure:

- a discriminant function is calculated using cross-validation for the existing X-ray spectral classes;
- eigenvectors from PCA on the COUP sample are used to calculate the principal components (PCs) for the additional X-ray sources;
- discriminant scores are then calculated for each additional X-ray source;
- additional X-ray sources are then assigned to an existing X-ray spectral class based on posterior probabilities.

Analysis of Other Star Formation Regions

A web-based archive called An Archive of *Chandra* Observations of Regions of Star Formation[1] (ANCHORS, Spitzbart et al. 2005) is being created by the Harvard-Smithsonian Center for Astrophysics. ANCHORS is being designed to aid both the X-ray astronomer in comparing X-ray datasets and the star formation astronomer in comparing stars across the spectrum (Spitzbart et al. 2005). This database will contain every point source observed during *Chandra* observations of star formation regions. Currently, the database contains X-ray source properties for ~50 *Chandra* fields totaling 10,000+ sources. We have been and will continue to work with the ANCHORS team to obtain star formation region datasets. As our application of the classification algorithm to the ANCHORS spectral database proceeds, the X-ray spectral classification results will be included in the ANCHORS database.

Preliminary Results For NGC 1333

We ran our extended algorithm on several star formation regions from the ANCHORS database, to ascertain whether the results from the Orion Nebula Cluster generalize to other

[1] http://hea-www.harvard.edu/~swolk/ANCHORS/

young stellar clusters. Preliminary results are presented here for the *Chandra* archival observation of NGC 1333, a highly active star formation region within the Perseus molecular cloud complex noted for its large population of protostars and young stellar outflows (Getman et al. 2002). Source detection was performed by the ANCHORS team, using the same CIAO-based wavelet source detection program that we have used for previous *Chandra* archival observations. We then used standard CIAO tools and IDL to extract spectra and analyze data. Detected sources with fewer than 100 total counts were eliminated, to limit faint sources with poor photon counting statistics. This resulted in a subset of 28 out of 90 detected sources.

Figure 3 shows the resulting classification of 17 of the NGC 1333 sources combined with the original plot from the 444 COUP sources in Figure 1. These 17 sources have class membership probabilities greater than 0.90. The 17 NGC 1333 sources are enclosed within black boxes. Two NGC 1333 sources with nearly identical RA and DEC are enclosed within a dotted-line box in class 8. One of the two sources is a known, deeply-embedded, classical T Tauri star (Rodriguez et al. 1999). The other component is an X-ray discovered companion with a projected separation of 2" (Getman et al. 2002). Both components have nearly identical spectra with log N_H of 22.2, suggesting that they may constitute a physical binary (Getman et al. 2002). Source #22 is a young stellar object (YSO) in NGC 1333 that our algorithm placed in class 3. In Figure 3, it appears to be located closer to the centroid of class 2 rather than class 3. However, this plot only shows the first two PCs. Four PCs were used to classify the sources. When a 3-D plot of the horseshoe is viewed, it is clear that NGC 1333 source #22 is much closer to the centroid of class 3 than class 2. A progression of rotated plots of the first three PCs is shown in Figure 4. These plots display the horseshoe-shaped graph rotated about the y-axis (viewed from the left side of the graph in Figure 3).

Known optical properties of the 17 NGC 1333 sources (Getman et al. 2002) fall within the ranges given in Table 1, following the trends found for the ONC. From this data, it can be seen that the application of our algorithm to NGC 1333 sources demonstrates that the general classification results for Orion are followed closely by a second star formation region. The results from Orion and preliminary results from NGC 1333 indicate that our

non-parametric approach to X-ray spectral classification can yield basic information concerning the nature of X-ray emission from young stars. Specifically, nearly all PMS stars lie in a single spectral sequence, with the exception of the outlier class 17 (the Beehive Proplyd, which displays an extraordinary double-peaked X-ray spectrum; Kastner et al. 2005). In addition, stars with strong flares tend to fall in the lowest number classes while stars exhibiting more consistent but lower-level coronal activity tend to fall in the higher numbered classes.

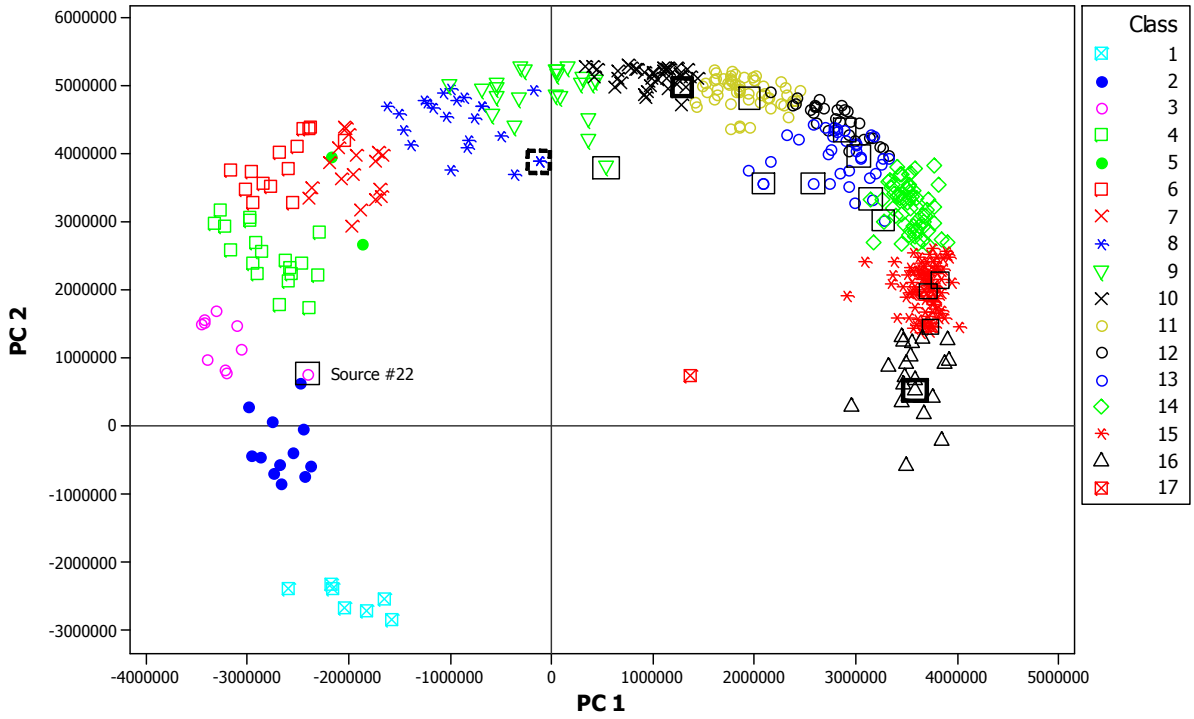

Figure 4: Plot of the first 2 principal components showing the NGC 1333 sources overlaid on the COUP horseshoe plot. Each of the 17 NGC 1333 sources is enclosed in a box.

Several other star formation regions (NGC 2264, IC 348, LYNDS 1551, and M8) exhibiting various environments, distances from our solar system, and *Chandra* observation exposure times, have been selected for additional testing and demonstration/determination of the ONIR trends found thus far. This testing and analysis will be completed in the second year of our project.

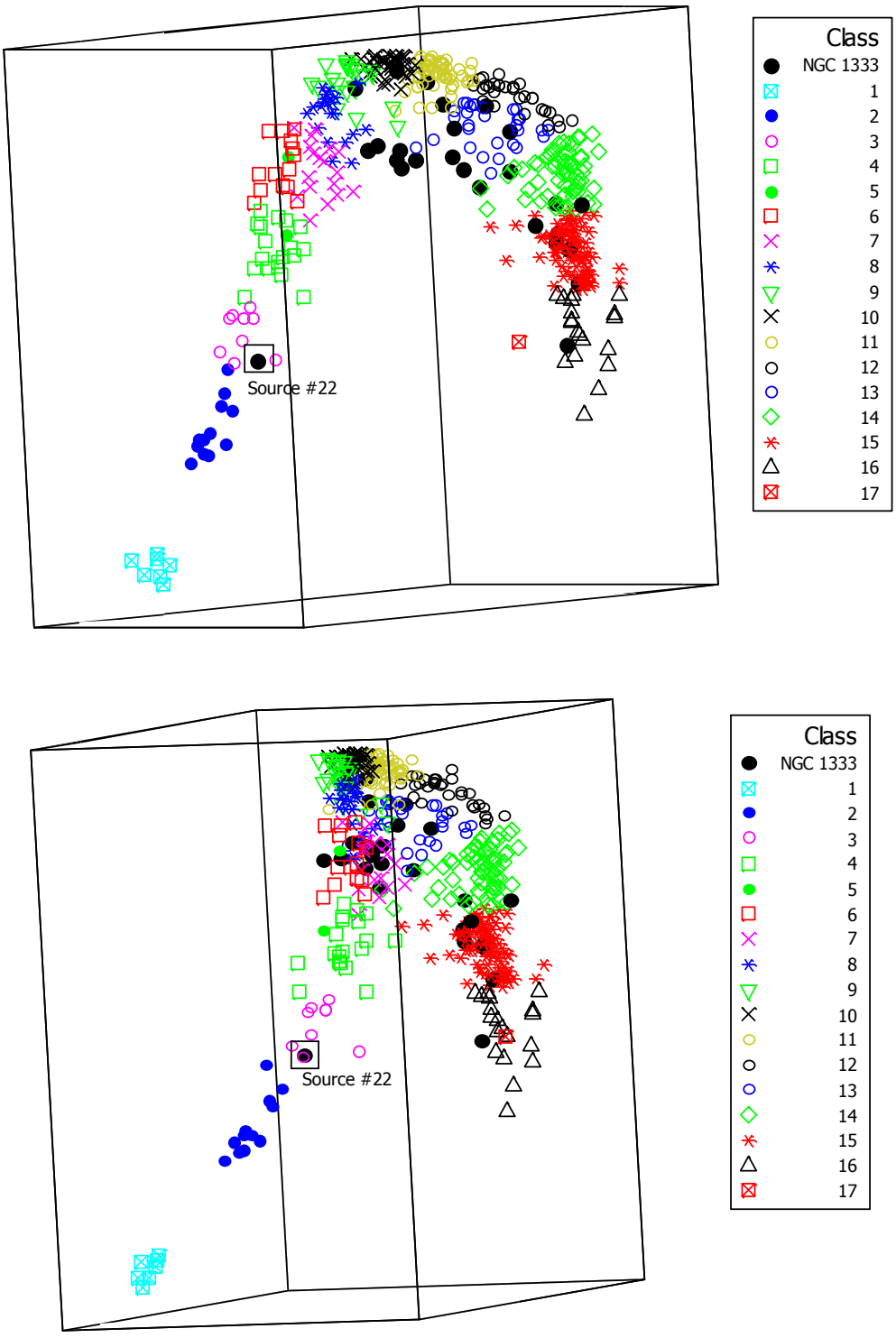

Figure 4: Rotation sequence of plot of the first 3 principal components showing location of NGC 1333 source #22 in three dimensions.

Addition of Source Variability Inputs to Algorithm

The study of variability of X-ray sources is a young field and an extremely active area of astrophysical research. We have begun work on adding X-ray source variability data to the classification algorithm. The first step is to investigate potential temporal descriptors with the purpose of identifying those that are intrinsic to sources of X-ray emission. A literature review was performed on light curve studies. In addition, discussions were held with astronomical researchers during the PI's AISRP-funded collaborative visit to the Palermo Observatory in Spring 2006. As a result, the following variability descriptors are currently under investigation, although more descriptors may be added, particularly from power spectrum analysis of the X-ray light curve.

Period: the X-ray period of the source as determined from the X-ray light curve (Flaccomio et al. 2005).

The X-ray period is difficult to determine due to the variable nature of young stars. A long exposure time is needed to accurately determine this value. Although the exposure time for the COUP observation was ~850 ksec, a typical Chandra observation has an exposure time between 10 and 100 ksecs. The X-ray period cannot be reliably determined for most of the observations of star formation regions currently in the *Chandra* archive, due to their relatively short exposure times (in comparison to the long exposure time of the COUP observation). Therefore, this descriptor has been eliminated from our study, since there currently are not enough observations with sufficiently long exposure times to accurately determine the X-ray period of many PMS stars.

Characteristic Level: the level of X-ray emission between isolated flare events that is typical or representative of a particular X-ray source during the observation (Wolk et al. 2005).

The characteristic level is considered an intrinsic property of the source and can be used to discriminate different types of sources (i.e., impulsive sources versus constant sources).

This value was determined for each X-ray source in the COUP observation by using a maximum likelihood block (MLB) analysis of each X-ray light curve (Flaccomio et al. 2005). The MLB approach identifies blocks for which the X-ray luminosity of the source is above, below, or at the characteristic level.

A related measure is the fraction of time that the source is at its characteristic level. This measure represents the amount of time, during the observation, that the X-ray luminosity of the source is somewhat constant. This measure is highly dependent on the exposure time for the observation. Therefore, this descriptor was eliminated from our study.

Median: the median value from the cumulative distribution of photon counts for a particular source.

Normalized Range: the range of the light curve in photon counts per ksec, normalized to the characteristic level of the source. This is the maximum to minimum range of the entire light curve.

Maximum Flare: a measure of the largest flare in the light curve, measured from the characteristic level. This measure can be used to distinguish stars that are variable in X-rays from those that are not variable in X-rays.

Principal Component Analysis Using Temporal Variables

We are performing exploratory principal component analyses on several combinations of the X-ray variability descriptors. The preliminary results from running PCA on characteristic level, normalized range, and median are shown in Figure 5.

The Figure contains a dense cluster of sources at the center of the X-ray variability PC diagram --- which likely represents the “typical” X-ray flux temporal behavior of young stars --- and three sequences of outliers. These outlier sequences reveal sources that have

(a) unusually large median fluxes, as measured over a long duration, (b) unusually large characteristic (as opposed to median) X-ray levels, and (c) unusually large ranges of X-ray flux. Hence, this diagram can be used as a temporal diagnostic due to its potential to isolate two important classes of bright X-ray sources – those with large characteristic fluxes (class b) and those that exhibit strong flares (class c).

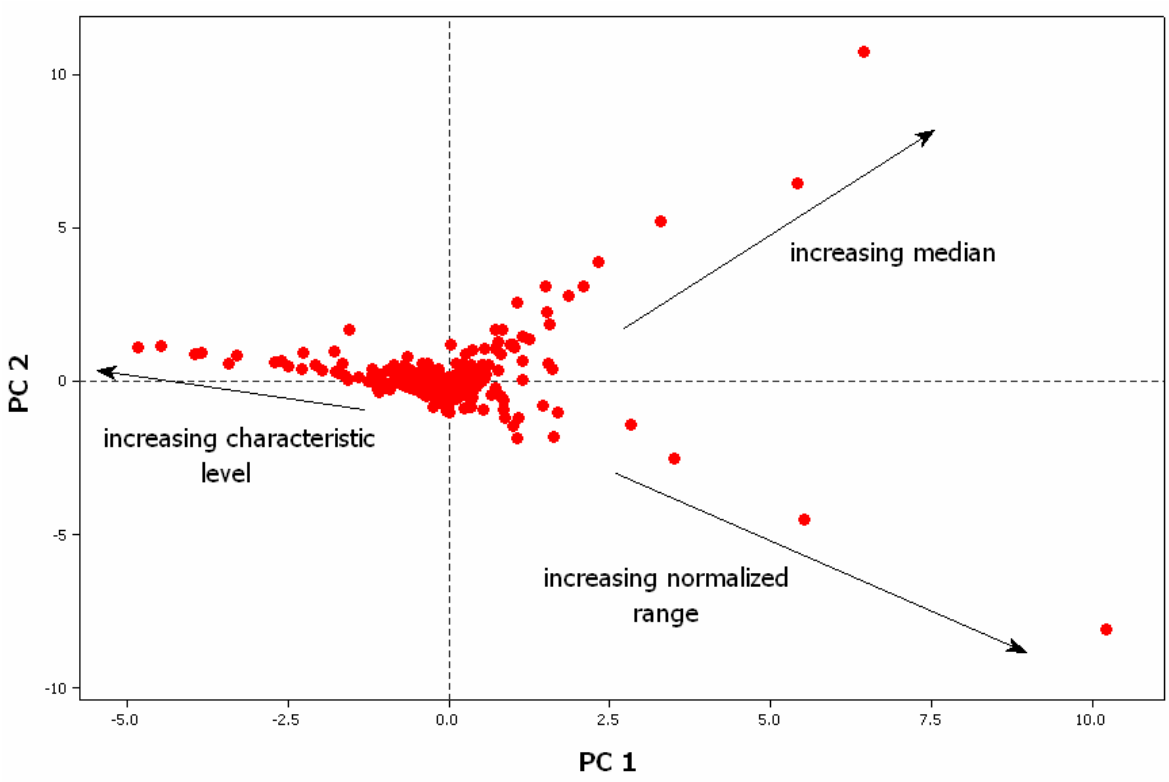

Figure 5: Principal component plot from PCA on the X-ray source variability descriptors of characteristic level, normalized range, and median.

Further analysis of the results of the X-ray variability modifications made to the algorithm will be presented in the next report.

Dissemination of Results

Several papers and conference presentations are being developed to summarize our results. The first paper, "An X-ray spectral classification algorithm with application to young

stellar clusters" (Hojnacki, S.M., Kastner, J.H., Micela, G., LaLonde, S.M., and Feigelson, E.D.) is in the final draft stage and will be submitted to the Astrophysical Journal in the September 2006 timeframe.

Following is a list of other papers in progress, to be submitted to refereed journals:

- Beyond Orion: Application of the X-ray spectral classification algorithm to other star formation regions
- Classification of X-ray sources using a temporal/spectral algorithm

REFERENCES

Flaccomio, E. et al. 2005, ApJS, 160, 450

Getman, K., et al. 2002, ApJ, 575, 354

Getman, K., et al. 2005, ApJS, 160, 319

Hojnacki, S. M., & Kastner, J. H. 2004, Proceedings of the SPIE, 5493, 474

Kastner, J. H., et al. 2005, ApJS, 160, 511

Rodriguez, L.F., et al. 1999, ApJS, 125, 427

Spitzbart, B., et al. 2005, "Proc of Star Form in the Era of Three Great Observatories"

Wolk, S. et al. 2005, ApJS, 160, 423